Stochastic Analysis of Unsaturated Flow with Probabilistic Collocation Method

Weixuan Li[a], Zhiming Lu[b], and Dongxiao Zhang[a,c]

[a] Department of Energy and Resource Engineering, College of Engineering, Peking University, Beijing, China;

[b] Hydrology and Geology Group (EES-6), Los Alamos National Laboratory, Los Alamos, New Mexico, USA;

[c] Sonny Astani Department of Civil and Environmental Engineering, and Monk Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, California, USA

First name	Last name	Telephone number	Email address
Weixuan	Li	86 10 62762547	weixuan@pku.edu.cn
Zhiming	Lu	505 6652126	zhiming@lanl.gov
Dongxiao	Zhang	213 7400567	donzhang@usc.edu

Abstract

In this study, we present an efficient method, called the probabilistic collocation method (PCM), for uncertainty analysis of flow in unsaturated zone, in which the constitutive relationship between the pressure head and the unsaturated conductivity is assumed to follow the van Genuchten-Mualem model. Spatial variability of soil parameters in the von Genuchten model leads to uncertainty in predicting flow behaviors. The aim is to quantify the uncertainty associated with flow quantities such as the pressure head and the effective saturation. In the proposed approach, the input random fields, i.e., the soil parameters are represented via the Karhunen-Loeve expansion and flow quantities are expressed by polynomial chaos expansions (PCEs). The coefficients in the PCEs are determined by solving the equations for a set of carefully selected collocation points. To illustrate this approach, we use two-dimensional examples with different input variances and correlation scales. We also demonstrate how to deal with multiple input random parameters, including the uncorrelated and perfectly correlated cases. To validate the (PCM), we compare the mean and variance of the pressure head derived from the PCM and those from Monte Carlo (MC) simulations. The comparison reveals that the PCM can accurately estimate flow statistics with a much smaller computational effort than the MC.

Index terms: 1805 Computational hydrology; 1869 Stochastic hydrology; 1873 Uncertainty assessment; 1875 Vadose zone

1. Introduction

One of the crucial problems in modeling flow and transport in the subsurface is the treatment of uncertainty. Uncertainty may be caused by a number of factors. It is well known that geological media exhibit a high degree of spatial variation over various scales. The properties that control flow and transport in the media, such as permeability and porosity, are also strongly heterogeneous in space. This spatial variability may have a strong impact on fluid flow in the media. Furthermore, these properties are usually measured only at a limited number of locations because of the high cost associated with subsurface measurements. Although media properties are deterministic, due to the lack of information it is common to treat them as spatially varying random fields, characterized by statistic moments that are derived from a limited number of measurements. In turn, the partial differential equations governing the subsurface flow in such media become stochastic.

In this study, we consider steady-state flow in the heterogeneous vadose zone, which connects the hydrology process above the land surface and the saturated aquifer in the subsurface. The vadose zone also acts as a buffer and passage in the process of pollutants movement from the land surface to the groundwater. Because of its important role in determining the pathway of pollutants, the vadose zone has received increasing attention in recent years. Because of the coexistence of water and air phases in this zone, the equation governing the flow in this zone becomes nonlinear, i.e., the hydraulic conductivity depends on the pressure head. The nonlinear property coupled with uncertainty leads to a great complexity in the numerical simulations.

Many stochastic approaches have been developed to study the effects of spatial variability on flow in saturated and unsaturated zone (Jury, 1982; Yeh et al.,1985a,b; Mantoglou and Gelhar, 1987; Mantoglou,1992; Russo, 1993, 1995; Zhang and Winter, 1998; Zhang, 1998, 1999, 2002; Lu and Zhang, 2002; Yang et al, 2004). The Monte Carlo (MC) simulation is the best-known and widely used approach in solving stochastic equations. As a statistical sampling approach, the MC is conceptually straightforward and easy to implement. The input random parameters are sampled repeatedly and independently from prescribed distributions, which may be obtained based on the field observations. Then, for each realization (sample) of input random fields, deterministic governing equations are solved to obtain the corresponding realization of output random fields. The required statistical properties, such as the statistical moments and probability density functions, can then be estimated based on these output realizations. A large number of realizations are needed to achieve statistical convergence, depending on the variability of the input parameters. Such a procedure usually leads to a high computational cost. As such, the applicability of MC is often limited to small-scale problems. In this study, for the purpose of validating the proposed approach, a large number of MC simulations are used and the results from these MC simulations are considered as the reference.

In this study, a Karhunen-Loeve (KL) expansion based probabilistic collocation method (PCM) is presented for predicting flow in the vadose zone. This approach has been used for stochastic analysis in some fields (Webster et al., 1996; Tatang et al., 1997). Coupled with the Karhunen-Loeve expansion of the random permeability field, Li and Zhang (2007) applied the PCM method to the simulation of flow in saturated heterogeneous porous media. In this

85 approach, the input random field is first expressed as the sum of its mean field and a zero
86 mean perturbation, which is further decomposed into a KL expansion with an infinite number
87 of terms. By truncating the KL series at a finite number of terms, the stochastic model is
88 simplified into finite stochastic dimensions. That is, the random field is represented with a
89 finite set of independent random variables. The steps in implementing the PCM are similar to
90 those of MC in that replicates of the random field are solved deterministically. However, the
91 input replicates are not randomly and equal-probably sampled but selected following certain
92 rules and thus referred to as "representations" in this work. The objective of these selection
93 rules is to significantly reduce the number of model simulations required for adequate
94 estimation of output uncertainties, compared to the conventional MC method.
95

96 ## 2. Stochastic differential equations

97 Consider steady state flow in unsaturated porous media satisfying the following continuity
98 equation and Darcy's law:
99

100 $$\nabla \cdot \mathbf{q}(\mathbf{x}) = g(\mathbf{x}), \tag{1}$$

101

102 $$\mathbf{q}(\mathbf{x}) = -K(\mathbf{x},\psi)\nabla[\psi(\mathbf{x}) + z], \tag{2}$$

103

104 subject to the following boundary conditions:
105

106 $$\psi(\mathbf{x}) = H_B(\mathbf{x}), \qquad \mathbf{x} \in \Gamma_D, \tag{3}$$

107

$$\mathbf{q}(\mathbf{x}) \cdot \mathbf{n} = Q(\mathbf{x}), \qquad \mathbf{x} \in \Gamma_N , \tag{4}$$

where $\mathbf{q}$ is the specific discharge (flux), g is the sink/source term, ψ is the pressure head, z is the elevation, $\psi + z$ is the total head, and $K(\mathbf{x}, \psi)$ is the unsaturated hydraulic conductivity, which depends on pressure head ψ. $H_B(\mathbf{x})$ and $Q(\mathbf{x})$ are prescribed pressure head and flux on Dirichlet and Neumann boundary segments, respectively.

To solve this set of equations described above, one must specify the constitutive relationship between K and ψ. Some empirical models have been investigated, including the Gardner-Russo model (Gardner, 1958; Russo, 1988), the Brooks-Corey model (Brooks and Corey, 1964), and van Genuchten-Mualem model (van Genuchten, 1980). In this study, we adopt the van Genuchten-Mualem model:

$$K(\mathbf{x}) = K_s(\mathbf{x}) \sqrt{S_e(\mathbf{x})} \{1 - [1 - S_e^{1/m}(\mathbf{x})]^m\}^2 , \tag{5}$$

$$S_e(\mathbf{x}) = \{1 + [-\alpha(\mathbf{x})\psi(\mathbf{x})]^n\}^{-m} , \tag{6}$$

where K_s is the saturated hydraulic conductivity, $S_e = \theta_e / (\theta_s - \theta_r)$ is the effective saturation, θ_e is the effective water content, θ_s and θ_r are the respective saturated and residual water content, α is a fitting parameter that is inversely related with the pore size distribution, n is another fitting parameter, and $m = 1 - 1/n$. The dependent variables ψ and S_e can be written as functions of space coordinate ($\mathbf{x}$), sink/source (g), boundary conditions (H_B, Q), and soil properties ($K_s, \alpha, n, \theta_s, \theta_r$):

$$\psi = \psi(\mathbf{x}, g, H_B, Q, K_s, \alpha, n, \theta_s, \theta_r), \quad (7)$$

$$S_e = S_e(\mathbf{x}, g, H_B, Q, K_s, \alpha, n, \theta_s, \theta_r). \quad (8)$$

Uncertainty associated with any argument in ψ and S_e may lead to uncertainty of ψ and S_e. In this study, we assume K_s, α, n are random fields whereas other arguments are deterministic. Our purpose is to estimate the statistical properties, i.e., the mean and variance, of the flow quantities such as the pressure and effective saturation, which are the output fields in our model.

3. Karhunen-Loeve Expansion

The first step of solving stochastic equations is to find a proper way to represent the input and output random fields. In this study, we use Karhunen-Loeve expansion (KL) to represent the input fields, for given covariance functions of the input fields. Since the covariance structures of the output fields are not known in advance, they cannot be expanded using the KL expansion. Instead, they are expressed in a form of Polynomial Chaos Expansion (PCE).

Consider an input random field $U(\mathbf{x}, \omega)$, where $\mathbf{x} \in D$ is the coordinates in the physical domain and $\omega \in \Omega$ denotes the coordinates in the probability space. It is assumed that the mean and covariance function of $U(\mathbf{x}, \omega)$ are known: $\bar{U}(\mathbf{x}) = < U(\mathbf{x}, \omega) >$, $C(\mathbf{x}_1, \mathbf{x}_2) = < [U(\mathbf{x}_1, \omega) - \bar{U}(\mathbf{x}_1)][U(\mathbf{x}_2, \omega) - \bar{U}(\mathbf{x}_2)] >$. These statistic moments can be estimated from the field data. For example, for any two points $\mathbf{x}_1 = (x_{11}, x_{12}, x_{13})^T$ and $\mathbf{x}_2 = (x_{21}, x_{22}, x_{23})^T$ in domain D, the covariance function may take the separate exponential form

$$C(\mathbf{x}_1, \mathbf{x}_2) = \sigma_U^2 \exp[-\sum_{i=1}^{3} \frac{|x_{1i} - x_{2i}|}{\lambda_i}], \quad (9)$$

or the Gaussian form

$$C(\mathbf{x}_1, \mathbf{x}_2) = \sigma_U^2 \exp[-\frac{\pi}{4} \sum_{i=1}^{3} (\frac{x_{1i} - x_{2i}}{\lambda_i})^2]. \quad (10)$$

In the above, σ_U^2 is the variance, λ_i is the correlation length in the i^{th} dimension. By definition, the covariance function is symmetric and positive definite, which means that it can be decomposed as (Courant and Hilbert, 1953):

$$C(\mathbf{x}_1, \mathbf{x}_2) = \sum_{i=1}^{\infty} \eta_i U_i(\mathbf{x}_1) U_i(\mathbf{x}_2), \quad (11)$$

where η_i and $U_i(\mathbf{x})$ are the eigenvalues and eigenfunctions of the covariance function, respectively. They can be solved from the following Fredholm equation of second kind:

$$\int_D C(\mathbf{x}_1, \mathbf{x}_2) U_i(\mathbf{x}_2) d\mathbf{x}_2 = \eta_i U_i(\mathbf{x}_1). \quad (12)$$

Because of the symmetry and the positive definiteness of the covariance function, its eigenvalues are positive and real, and its eigenfunctions are orthogonal and form a complete set,

177 $$\int_D U_i(\mathbf{x})U_j(\mathbf{x})d\mathbf{x} = \delta_{ij}\,, \tag{13}$$

178

179 where δ_{ij} is the Kronecker delta function, which equals to one for $i = j$ and zero otherwise.

180 Then the random field can be expressed as

181

182 $$U(\mathbf{x},\omega) = \bar{U}(\mathbf{x}) + U'(\mathbf{x},\omega) = \bar{U}(\mathbf{x}) + \sum_{i=1}^{\infty} \xi_i(\omega)\sqrt{\eta_i}U_i(\mathbf{x})\,, \tag{14}$$

183

184 where ξ_i are a set of orthogonal random variables satisfying $<\xi_i>=0$ and $<\xi_i\xi_j>=\delta_{ij}$.

185 When the underlying random field is Gaussian, ξ_i are standard Gaussian random variables.

186 The expansion in equation (14) is called the Karhunen-Leove (KL) expansion. The random

187 field $U(\mathbf{x},\omega)$ is decomposed as the sum of its mean and a perturbation, which can be further

188 represented by a series of KL terms. Without loss of generality, it is assumed that the

189 eigenvalues have been sorted such that $\eta_1 \geq \eta_2 \geq ...$ and their corresponding eigenfunctions

190 are also arranged accordingly. By truncating the infinite KL series at the Nth term, $U(\mathbf{x},\omega)$ is

191 approximated via N independent random variables ξ_i, $i = 1,...,N$, weighted by the

192 eigenvalues and deterministic eigenfunctions. When the underlying random field is Gaussian,

193 this approximation is optimal with mean square convergence. For some particular covariance

194 functions defining on regular domains (such as rectangular domains in 2D), eigenvalues and

195 eigenfunctions can be solved analytically (Ghanem and Spanos, 1991; Zhang and Lu, 2004).

196 However, in general, the integral equation, i.e., (12), has to be solved numerically (Ghanem

197 and Spanos, 1991).

198

199 One of the interesting features of the KL expansion is that the sum of all eigenvalues is

200 related to the total variability of the input field. Setting $\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}$ in equation (11),
201 integrating it over the domain D, and recalling orthogonality of eigenfunctions yields
202

$$\begin{aligned}\int_D \sigma_U^2(\mathbf{x})d\mathbf{x} &= \int_D C(\mathbf{x},\mathbf{x})d\mathbf{x} \\ &= \int_D [\sum_{i=1}^{\infty} \eta_i U_i^{\,2}(\mathbf{x})]d\mathbf{x}, \\ &= \sum_{i=1}^{\infty} \eta_i\end{aligned} \tag{15}$$

203
204
205 where $\sigma_U^2(\mathbf{x}) = C(\mathbf{x},\mathbf{x})$ is the variance function of $U(\mathbf{x},\omega)$. If $U(\mathbf{x},\omega)$ is stationary, (15)
206 leads to $\sum_{i=1}^{\infty} \eta_i = |D|\sigma_U^2$, where $|D|$ is the measure of the domain D. Equation (15)
207 indicates that the total variability of $U(\mathbf{x},\omega)$ over the whole domain is distributed to all KL
208 terms, with the weight of η_i. The KL decomposition is a spectral decomposition. As will be
209 shown in the illustrative examples, different KL terms reflect the variability on different
210 length scales. So the physical meaning of the KL expansion is to separate the uncertainty on
211 different spatial scales: those terms corresponding to large eigenvalues (leading terms)
212 represent variability on larger scales and the terms corresponding to small eigenvalues for
213 variability on smaller scales. Thus we can effectively approximate the stochastic property of a
214 random field with relatively few random variables, by retaining those leading KL terms
215 (terms with large eigenvalues).
216

217 4. Polynomial Chaos Expansion

218 Because the output random field ψ or S_e depends on the input, it can be shown that ψ
219 or S_e is a function of the random vector $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_N)^T$, where ξ_i are the random
220 variables used to approximate the input parameters. However, the specific relationship

between the output random fields and $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_N)^T$ is yet to be determined.

Since the statistics of the output random fields are not known in advance, ψ or S_e cannot be decomposed using the KL expansion. Alternatively, the PCE is a more general representation that can be used under this condition. In the following derivations, we take the pressure head as an example. The effective saturation can be expanded in a similar way. Suppose ψ can be expanded by the following form:

$$\begin{aligned}\psi(\mathbf{x},\boldsymbol{\xi}) = a_0(\mathbf{x}) + \sum_{i_1=1}^{\infty} a_{i_1}(\mathbf{x})\Gamma_1(\xi_{i_1}) + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1} a_{i_1 i_2}(\mathbf{x})\Gamma_2(\xi_{i_1},\xi_{i_2}) \\ + \sum_{i_1=1}^{\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}(\mathbf{x})\Gamma_3(\xi_{i_1},\xi_{i_2},\xi_{i_3}) + ...\end{aligned}, \tag{16}$$

where $a_0(\mathbf{x})$ and $a_{i_1 i_2 ... i_d}(\mathbf{x})$ are deterministic coefficients. The basis $\Gamma_d(\xi_{i_1}, ..., \xi_{id})$ is a set of polynomial chaos with respect to those independent random variables $\xi_{i_1}, ..., \xi_{id}$ (Winener, 1938). For independent standard Gaussian random variables, $\Gamma_d(\xi_{i_1}, ..., \xi_{id})$ are the multidimensional Hermit polynomials with order of d. They are expressed as

$$\Gamma_d(\xi_{i_1}, ..., \xi_{i_d}) = (-1)^d e^{\frac{1}{2}\xi^T\xi} \frac{\partial^d}{\partial\xi_{i_1}...\partial\xi_{i_d}}(e^{\frac{1}{2}\xi^T\xi}), \tag{17}$$

where $\boldsymbol{\xi} = (\xi_{i_1}, ..., \xi_{id})^T$. By truncating the polynomial series in equation (16) at a certain order, we have an approximation of the output random field. In particular, the second-order approximation with Hermit polynomials can be written as

$$\hat{\psi}(\mathbf{x},\boldsymbol{\xi}) = a_0(\mathbf{x}) + \sum_{i=1}^{N} a_i(\mathbf{x})\xi_i + \sum_{i=1}^{N} a_{ii}(\mathbf{x})(\xi_i^2 - 1) + \sum_{i=1}^{N}\sum_{j=1}^{i-1} a_{ij}(\mathbf{x})\xi_i\xi_j \,, \tag{18}$$

or a simplified form

$$\hat{\psi}(\mathbf{x},\boldsymbol{\xi}) = \sum_{j=1}^{P} c_j(\mathbf{x})\Phi_j(\boldsymbol{\xi}) \,. \tag{19}$$

There is a one-to-one correspondence between the terms in (18) and (19). The total number of PCE terms is $P = (N+d)!/(N!d!)$, where N is the random dimensionality (the number of KL terms retained to represent the mean-removed random input fields) and d is the order of polynomial chaos.

One of the important properties of the polynomial chaos is that all the polynomials in equation (19) are mutually orthogonal, which means $<\Phi_i(\boldsymbol{\xi})\Phi_j(\boldsymbol{\xi})>=0$ for $i \neq j$, and for $i = j$, $<\Phi_j^2>$ can be evaluated easily (Ghanem and Spanos, 1991). Once the coefficients $c_j(\mathbf{x})$ are determined, as described in the next section, the statistical properties of the pressure head can be easily estimated from equation (19). For example, the mean pressure head is

$$\bar{\hat{\psi}}(\mathbf{x},\boldsymbol{\xi}) = <\hat{\psi}(\mathbf{x},\boldsymbol{\xi})> = \sum_{j=1}^{P} c_j(\mathbf{x}) <\Phi_j(\boldsymbol{\xi})> = c_1(\mathbf{x}) \,. \tag{20}$$

Note that in deriving equation (20), we have used the following properties of the polynomial chaos: $\Phi_1(\boldsymbol{\xi}) = 1$ and $<\Phi_j(\boldsymbol{\xi})> = <\Phi_j(\boldsymbol{\xi})\Phi_1(\boldsymbol{\xi})> = 0$ for $j = 2,3,...P$. The variance of the

pressure head is

$$\begin{aligned}\sigma_{\hat{\psi}}^2 &= <(\hat{\psi}(\mathbf{x},\xi) - <\hat{\psi}(\mathbf{x},\xi)>)^2> \\ &= \sum_{j=2}^{P}\sum_{k=2}^{P} c_j(\mathbf{x})c_k(\mathbf{x}) <\Phi_j(\xi)\Phi_k(\xi)>, \\ &= \sum_{j=2}^{P} c_j{}^2(\mathbf{x}) <\Phi_j^2>\end{aligned} \tag{21}$$

where $<\Phi_j^2>$ can be evaluated in advance. Higher-order terms can be calculated similarly.

5. Probabilistic Collocation Method

Galerkin's approach and Probabilistic Collocation Method (PCM) are two of the methods that may be used to determine the coefficients $c_j(\mathbf{x})$ in the polynomial expansion. Here we choose PCM, which leads to independent equations and is thus capable of easily dealing with complex non-linear problems.

5.1 Implementation of Probabilistic Collocation Method

The probabilistic collocation method has no essential differences compared with the typical collocation method except that the typical collocation method is used to seek a deterministic solution whereas the PCM is used to seek a random solution defined on probability space. Again, we only demonstrate how to determine the coefficients in the PCE approximation of the pressure head. The coefficients in the PCE approximation of the effective saturation can be determined similarly. Let $\psi = \psi(\mathbf{x},\xi)$ denote the solution of the nonlinear stochastic differential equations. We seek an approximate solution as described in

equation (19). We define the residual between the unknown solution $\psi(\mathbf{x},\boldsymbol{\xi})$ and this approximation as

$$R_P(\mathbf{x},\boldsymbol{\xi}) = \psi(\mathbf{x},\boldsymbol{\xi}) - \hat{\psi}(\mathbf{x},\boldsymbol{\xi}) . \tag{22}$$

The collocation method proceeds by requiring this residual vanishes at some sets of collocation points $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_P$. The procedure is summarized as follows (Li and Zhang, 2007):

First, choose P collocation points in the probability space, $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, ..., \xi_{iN})^T$, $i = 1, 2, ..., P$, as described in section 5.2. Secondly, substitute $\boldsymbol{\xi}_i = (\xi_{i1}, \xi_{i2}, ..., \xi_{iN})^T$ into the KL expansion to generate a representation of the input field. For each of the P representations, the differential equations are deterministic and can be solved to give an output $\psi(\mathbf{x},\boldsymbol{\xi}_i)$. This leads to P sets of independent equations that can be solved in parallel or in sequence. With the P sets of solutions and by letting $R_P(\mathbf{x},\boldsymbol{\xi}_i) = 0$, one has

$$\sum_{j=1}^{P} c_j(\mathbf{x})\Phi_j(\boldsymbol{\xi}_i) = \psi(\mathbf{x},\boldsymbol{\xi}_i), \quad i = 1, 2, ..., P . \tag{23}$$

For each location **x** in the simulation domain, the above equations form P linear equations for P unknowns $c_j(\mathbf{x})$. The matrix of coefficients $\Phi_j(\boldsymbol{\xi}_i)$ and the right-hand sides $\psi(\mathbf{x},\boldsymbol{\xi}_i)$ are known. By solving this linear system, the coefficients in the PCE expansion can be determined for locations of interest.

307

308 An alternate approach for determining the unknown coefficients is the Galerkin approach
309 (Ghanem and Spanos, 1991). In this approach, the residual is required to be orthogonal to the
310 basis functions $\Phi_j(\boldsymbol{\xi})$ that are used in the PCE expansion,

311

312 $$< R_P(\mathbf{x},\boldsymbol{\xi})\Phi_j(\boldsymbol{\xi}) >= 0, \qquad j = 1,2,...,P\,. \tag{24}$$

313

314 Again, we have P constraints to determine the P unknown coefficients $c_j(\mathbf{x})$. Note that
315 equation (24) results in *P* coupled equations. Solving these coupled equations could be very
316 complicated and time-consuming, especially when one considers spatially varying
317 coefficients (when the solution is not a random variable but a random field) or nonlinear
318 problems. On the other hand, the PCM method leads to uncoupled equations, each of which
319 can be solved with existing, deterministic simulators. This feature makes the PCM applicable
320 to linear or non-linear problems in a straightforward manner.

321

322 **5.2 Selection of the Collocation Points**

323 One key issue in the PCM approach is the selection of collocation points. Previous studies
324 have demonstrated that for a given order of the PCE approximation the coordinates of each
325 collocation point should be selected from the roots of the next higher order orthogonal
326 polynomial. This technique, which is analogous to Gaussian quadrature, will yield the result
327 that is much more accurate than the randomly selected samples (Webster et al., 1996; Tatang
328 et al., 1997). For the case of second-order Hermit PCE, the coordinates of collocation points
329 are selected from the roots of the third-order Hermit polynomial $\xi^3 - 3\xi$, i.e., $-\sqrt{3}, 0$,

330 and $\sqrt{3}$. Then each collocation point is a combination of these three roots; two examples of
331 the collocation points are $\boldsymbol{\xi}_1 = (\xi_{11}, \xi_{12}, ..., \xi_{1N})^T = (0, 0, ..., 0)^T$ and
332 $\boldsymbol{\xi}_2 = (\xi_{21}, \xi_{22}, ..., \xi_{2N})^T = (\sqrt{3}, 0, ..., 0)^T$. .
333
334 The number of collocation points required is equal to the number of coefficients to be
335 determined $P = (N+d)!/(N!d!)$, where N is the dimensionality of the probability space
336 and d is the order of PCE approximation (2 in this case). However, the number of available
337 points satisfying the preceding requirement, that is the number of different combinations of
338 $(d+1)$ roots, is $M = (d+1)^N$. Note that M is always larger than P, the number of points
339 needed. As a result, we may select *P* best combinations out of *M* combinations in total. One
340 may optimize the selection of the collocation points with the following consideration (Li and
341 Zhang, 2007).
342
343 1) Keep as many points as possible in the region with a high probability density. Actually
344 the probability density $\rho(\boldsymbol{\xi})$ is the weight in the integral for calculating the statistical
345 moments.
346
347 $$\bar{\psi}(\mathbf{x}) = \int_{\Omega} \psi(\mathbf{x}, \boldsymbol{\xi}) \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{25}$$
348
349 $$\sigma_{\psi}^2(\mathbf{x}) = \int_{\Omega} (\psi(\mathbf{x}, \boldsymbol{\xi}) - < \psi(\mathbf{x}, \boldsymbol{\xi}) >)^2 \rho(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{26}$$
350
351 By setting $\hat{\psi}(\mathbf{x}, \boldsymbol{\xi}) = \psi(\mathbf{x}, \boldsymbol{\xi})$ at the region with a high probability density one can
352 increase the accuracy of the estimated statistical moments. For instance, when
353 $\xi_1, \xi_2, ..., \xi_N$ are N independent standard Gaussian random variables, the random

point $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_N)^T$ follows the following joint distribution:

$$\rho(\boldsymbol{\xi}) = (2\pi)^{N/2} \exp(-\frac{\boldsymbol{\xi}^T \boldsymbol{\xi}}{2}), \tag{27}$$

and this density function reaches the highest value at the origin point $\boldsymbol{\xi} = (0, 0, ..., 0)^T$, which corresponds to the mean parameter field. Therefore, this point is always kept. In addition, from (27) it is seen that in selecting these points one should keep as many zeros as possible in $\boldsymbol{\xi}$.

2) The matrix with its components $\Phi_j(\boldsymbol{\xi}_i)$ in equation (23) must have a full rank. Thus the equations are mutually independent and a unique solution can be solved. To achieve this objective, one should first sort the available points in an order of decreasing probability density. For instance, the point $(0, 0, ..., 0)$ should be the first collocation point. Then, for the candidate of the $(i+1)^{th}$ collocation point, the $(i+1)^{th}$ row of matrix $\Phi_j(\boldsymbol{\xi}_i)$ must be linearly independent with the previous i rows. Otherwise, this candidate is abandoned and the point with the next highest probability density should be tested. For given (N,d), the sets of collocation points may be selected once and tabulated for other simulations.

If both ψ and S_e are represented via the same basis polynomial chaos, the collocation points for computing these two output fields can be the same, which means the computational effort for solving the statistical moments of the two output fields is almost the same compared with the computational effort for only one output field. For example, if both ψ and

377 S_e are represented via the second-order Hermit polynomial chaos and the input random
378 dimensionality is 10, each of the output field will have $P=(10+2)!/(10!2!)=66$
379 coefficients to be determined. We just chose 66 collocation points from the combinations of
380 the roots of the third-order Hermit polynomial and solve the deterministic equations at these
381 points, for pressure head and water content simultaneously, rather than select $2P$
382 collocation points and solve the deterministic equations $2P$ times, P for pressure and P
383 for water content.
384

385 ## 6. Illustrative examples

386 We consider steady state unsaturated flow in a two-dimensional vertical cross section of
387 size 3 by 3 meters, discretized into 30×60 rectangular elements of 0.1 by 0.05 meters. We
388 assign a constant deterministic flux rate $Q=-0.005m/d$ at the upper boundary (negative
389 value standing for infiltration), a constant deterministic pressure head $\psi=0$ (water table) at
390 the bottom, and no flow at the left and right boundaries. The saturated and residual water
391 content are $\theta_s=0.4$ and $\theta_r=0.01$, respectively.
392

393 Note that both K_s and α are positive quantities and n is always larger than one. Here
394 we assume that the log saturated hydraulic conductivity $f=\ln(K_s)$, the log pore size
395 distribution parameter $\beta=\ln(\alpha)$ and the fitting parameter $s=\ln(n-1)$ are second-order
396 stationary Gaussian random fields with mean $<U>$ and separate exponential covariance
397 $C_U(\mathbf{x}_1,\mathbf{x}_2)=\sigma_U^2\exp[-\sum_{i=1}^{2}\frac{|x_{1i}-x_{2i}|}{\lambda_{Ui}}]$, where $U=f,\beta,s$; σ_U^2 is the variance, λ_U is the
398 correlation length of U. The subscripts $i=1,2$ refer to the horizontal and vertical

dimensions, respectively. The variability of a parameter can also be given in terms of the coefficient of variation $CV_V = \sigma_V / < V >$, for $V = K_s, \alpha, (n-1)$. The mean values of the three input parameters are set to be $< K_s >= 1 m/d$, $< \alpha >= 2 m^{-1}$, and $< n >= 1.4$. With the known mean and the coefficient of variation of a log-normal random field V, the moments of the corresponding normal random field $U = \ln V$ can be easily calculated via following relations (e.g., Zhang, 2002):

$$< U >= 2 \ln < V > -0.5 \ln[< V >^2 (1 + CV_V^2)], \tag{28}$$

$$\sigma_U^2 = \ln[1 + CV_V^2], \tag{29}$$

or conversely,

$$< V >= \exp[< U > +0.5\sigma_U^2], \tag{30}$$

$$\sigma_V^2 = [\exp(\sigma_U^2) - 1] \exp[2 < U > +\sigma_U^2]. \tag{31}$$

We design a series of model cases with different model complexity (single input random field or multiple input random fields) and parameter variabilities. For each case, we first derive the statistical moments (mean and variance) of output fields using the Probabilistic Collocation Method (PCM). For the purpose of comparison, we conduct Monte Carlo (MC) simulations with a large number of realizations. For each single run of Monte Carlo or PCM simulations, the FEHM code (Zyvoloski et al., 1997) is used to solve the deterministic differential equations. Unless otherwise noted, the comparison of results from the MC

423 method and the PCM is illustrated only along the central vertical line (x=1.5m). In case 1, we
424 demonstrate calculated statistical moments of both the pressure head and the effective
425 saturation. In other cases, only the results of pressure head are shown for the sake of
426 succinctness.
427

428 **6.1 Single random input**

429 In the first four cases, we treat α and n to be deterministic but $f = \ln(K_s)$ as the only
430 input random field with different levels of variability and different correlation lengths as
431 shown in Table 1. These cases are designed to explore the impact of the input variability and
432 the correlation length on the accuracy and efficiency of the PCM. As described previously,
433 we decompose $f = \ln(K_s)$ using the KL expansion and retain the first N random terms:
434 $f(\mathbf{x},\omega) = \overline{f}(\mathbf{x}) + \sum_{i=1}^{N} \xi_{fi}(\omega)\sqrt{\eta_{fi}}\, f_i(\mathbf{x})$. The output random fields, the pressure head and the
435 water content, can be written as $\psi = \psi(\mathbf{x}, \xi_{f1}, \xi_{f2}, ..., \xi_{fN})$ and $S_e = S_e(\mathbf{x}, \xi_{f1}, \xi_{f2}, ..., \xi_{fN})$,
436 respectively, indicating that the dimensionality in the probability space is N.
437

438 The contours of some selected eigenfunctions in case 1 are plotted in Figure 1. It is shown
439 that the first eigenfunction represents the spatial variability on the large scale and the
440 subsequent eigenfunctions represent the spatial variability on smaller scales. The eigenvalues
441 and their summation for cases 1 and 2 are shown in Figure 2. A decaying trend of eigenvalues
442 can be observed, which suggests that more input variability is distributed on large spatial
443 scales. The decaying rate of eigenvalues depends on the correlation length relative to the
444 domain size. The eigenvalues in case 1 decay faster than those in case 2, where the
445 correlation scale is smaller. In the first three cases, we keep 20 terms in the KL expansion and

use the second-order PCE. Under these conditions, 231 (= 22!/20!/2!) times of simulations are needed. For comparison, we explore the convergence of MC simulations based on simulations up to 8000 realizations. In particular, we are interested in the accuracy of MC results from 231 realizations, which represent the more-or-less equivalent computational cost required by the PCM. The statistical moments computed from all 8000 MC realizations are considered to be the "true" solutions for assessing the accuracy of the PCM. The mean and variance of pressure head and effective saturation from case 1 are plotted in Figure 3. First, from the results, we observe that the PCM solutions are in good agreement with those from 8000 MC simulations. However, the MC results computed from 231 realizations deviate substantially from the MC results from 8000 realizations ("true" solutions). Since the computational efforts for 231 PCM simulations are more or less the same as those for 231 MC simulations, the comparison indicates that the PCM is computationally more efficient than the MC simulations. Actually, for the mean pressure head, about 2000 MC simulations are needed to obtain the convergent result in this example. While for the pressure head variance, about 4000 MC simulations are necessary to yield a convergent result (Fig. 4). Secondly, the pressure head variance from the PCM is symmetric with respect to the vertical central line (shown in Figure 5), which is consistent with the symmetric boundary conditions on the left and right boundaries. For the MC approach, the symmetry of the pressure variance can be achieved only when a large number of simulations are conducted. Furthermore, the profiles of the curves in Figure 3(a) and Figure 3(b) are quite similar to that in (c) and (d). This is because the effective saturation is directly dependent on the pressure head. At the upper part of the domain, a lower pressure head leads to a lower water content and a large variability in the water content is consistent with a large variability in the pressure head.

Figure 6 compares the pressure head variance derived from both the MC method and the PCM for different degrees of conductivity variability and two different correlation lengths. In cases 1 and 2, all input parameters are the same, except that the conductivity variability in case 2 is smaller than that in case 3. It is seen from the figure that the PCM is more accurate when the conductivity variability is small. In cases 1 and 3, all input parameters are the same, except that the correlation length of the input random field in case 1 is larger than that in case 3. The figure indicates that the accuracy of the PCM decreases with the decrease of the correlation length. The reason is that the accuracy of the truncated KL expansion depends on the ratio of the domain size and the correlation length. It has been shown that, for a small correlation length, more terms are needed in the truncated KL expansion to retain the same accuracy (Ghanem and Spanos, 1991; Zhang and Lu, 2004). We also run case 4 in which all the input parameters are the same as those in case 3 but the number of retained KL terms is 30. Under this condition, 496 simulations are required to implement the second-order PCM approach. It can be seen that when more terms are included in the truncated KL expansion, the results from the PCM are closer to the MC results.

6.2 Multiple input random fields

One of the advantages of the PCM is the ease in dealing with multiple input random parameters, especially for uncorrelated or perfectly correlated inputs. In this section, we demonstrate the implementation of the PCM with multiple inputs by the following two cases.

6.2.1 Uncorrelated input random fields

In case 5, we assume that all the three input parameters are mutually uncorrelated random

493 fields. With this assumption, we can extend the previous PCM procedure from single input to
494 multiple inputs simply by increasing the input random dimensionality. Express the three input
495 parameters in the form of the KL expansion separately:
496
497 $$f(\mathbf{x},\omega) = \overline{f}(\mathbf{x}) + \sum_{i=1}^{N_f} \xi_{fi}(\omega)\sqrt{\eta_{fi}}\, f_i(\mathbf{x}), \tag{32}$$
498
499 $$\beta(\mathbf{x},\omega) = \overline{\beta}(\mathbf{x}) + \sum_{i=1}^{N_\beta} \xi_{\beta i}(\omega)\sqrt{\eta_{\beta i}}\, \beta_i(\mathbf{x}), \tag{33}$$
500
501 $$s(\mathbf{x},\omega) = \overline{s}(\mathbf{x}) + \sum_{i=1}^{N_s} \xi_{si}(\omega)\sqrt{\eta_{si}}\, s_i(\mathbf{x}), \tag{34}$$
502
503 where $\{\xi_{fi}\}$, $\{\xi_{\beta i}\}$ and $\{\xi_{si}\}$ are all mutually independent Gaussian random variables. The
504 pressure head can be written as $\psi = \psi(\mathbf{x}, \xi_{f1},...,\xi_{fN_f}, \xi_{\beta 1},...,\xi_{\beta N_\beta}, \xi_{s1},...,\xi_{sN_s})$, indicating that
505 the total degree of freedom in the probability space is $N = N_f + N_\beta + N_s$.
506
507 We truncate the KL expansion of $f = \ln(K_s)$, $\beta = \ln(\alpha)$ and $s = \ln(n-1)$ by keeping 10,
508 15, 15 terms, respectively. The total degree of freedom is 40. The mean and variance of
509 pressure derived from both PCM and MC are compared along the central vertical profile, as
510 shown in Figure 7. It is shown that the PCM performs well. Compared with the single input
511 cases, due to high random dimensionality, the number of simulations required in the PCM is
512 increased significantly (from 231 to 861).
513
514 **6.2.2 Perfectly correlated input random fields**

515 In case 6, we assume that there is a positive perfect correlation between $f = \ln(K_s)$ and
516 $\beta = \ln(\alpha)$, but both of them are uncorrelated with $s = \ln(n-1)$. This assumption is based on
517 the following consideration. From its physical meaning, α is inversely related with the
518 bubbling pressure. On the other hand, both the bubbling pressure and the saturated
519 conductivity are directly dependent on the pore size. A large pore size not only causes a high
520 saturated conductivity but also suggests a low bubbling pressure and hence a large α. Some
521 previous experimental studies indicate such a positive correlation (Simunek et al., 1998).
522 Since there is no justifiable data to reveal a specific covariance function between f and β,
523 we just assume they are positively and perfectly correlated in our model.
524
525 A perfect correlation between $f = \ln(K_s)$ and $\beta = \ln(\alpha)$ means that there is a linear
526 relationship between them $f = a\beta + b$, where *a* and *b* are constants. From this relationship,
527 one has $\bar{f} = a\bar{\beta} + b$ and $\sigma_f = a\sigma_\beta$, which can be written as a normalized form:
528
529 $$\frac{f(x,\omega) - \bar{f}(x)}{\sigma_f} = \frac{\beta(x,\omega) - \bar{\beta}(x)}{\sigma_\beta} = Y(x,\omega), \tag{35}$$
530
531 where Y is a Gaussian field with zero mean and unit variance. We approximate Y using
532 the truncated KL expansion
533
534 $$Y(\mathbf{x},\omega) = \bar{Y}(\mathbf{x}) + \sum_{i=1}^{N_Y} \xi_{Yi}(\omega)\sqrt{\eta_{Yi}}Y_i(\mathbf{x}), \tag{36}$$
535
536 and represent the two perfectly correlated input random fields via $Y(\mathbf{x},\omega)$:
537

$$f(\mathbf{x},\omega) = \overline{f}(\mathbf{x}) + \sigma_f Y(\mathbf{x},\omega), \tag{37}$$

$$\beta(\mathbf{x},\omega) = \overline{\beta}(\mathbf{x}) + \sigma_\beta Y(\mathbf{x},\omega). \tag{38}$$

Similarly, we decompose s in a KL form:

$$s(\mathbf{x},\omega) = \overline{s}(\mathbf{x}) + \sum_{i=1}^{N_s} \xi_{si}(\omega)\sqrt{\eta_{\beta i}}\, s_i(\mathbf{x}), \tag{39}$$

where $\{\xi_{Yi}\}$ and $\{\xi_{si}\}$ are mutually independent. Thus the pressure head can be written as $\psi = \psi(\mathbf{x}, \xi_{Y1},...,\xi_{YN_Y}, \xi_{s1},...,\xi_{sN_s})$, indicating that the total degree of freedom in the probability space is $N = N_Y + N_s$. We keep 15 KL terms to represent both Y and s, respectively. The total degree of freedom is 30.

Again, the mean and variance of pressure derived from both the PCM and MC are compared along the central vertical profile. It is shown in Figure 8 that the results given by PCM and MC are in good agreement. However, the variances of pressure head given by both approaches are much less than that in the uncorrelated case. It may be explained qualitatively. On the basis of equations (5) and (6) in the van Genuchten model, at a certain pressure the effective saturation will decrease as α becomes larger (note that the pressure is negative). Thus the relative conductivity will become smaller. However if saturated conductivity increases simultaneously, some of their impacts on the conductivity $K(\mathbf{x},\psi)$ will be neutralized with each other. Therefore, the pressure variance is reduced.

7. Discussions and Conclusions

Although the stochastic equations describing flow in unsaturated zone is complex because of the nonlinearity, in this study we demonstrated that the Probability Collocation Method (PCM) is still applicable. Like the MC method, the PCM is based on solving a set of deterministic equations. The difference between the two approaches is that the PCM requires the solutions at a set of selected collocation points whereas the MC requires the solutions at random sampling points. Both approaches can be implemented straightforwardly with the availability of a deterministic simulator. Like the MC, the PCM can be applied to various problems, either linear or non-linear, either with single or multiple inputs.

Because the stochastic structures of both input and output random fields have been carefully considered, the PCM can capture the stochastic behavior of the dependent variables such as the pressure field and the effective saturation by a small number of model simulations. Hence the efficiency of the PCM is significantly increased compared to the MC. This advantage is crucial in solving large-scale problems because solving each deterministic equation may require a large computational effort.

As shown in the illustrative examples, the PCM performs better when the input correlation scale is relatively large and the input variance is relatively small. If the correlation scale is too small or the input variance is too large, the PCM may yield inaccurate results. Actually, the truncations in the KL and PCE approximations are two major sources of errors in the PCM procedures, and the accuracy of KL and PCE approximations depend on the input correlation length and input variance, respectively. Our on-going research attempts to derive posterior error estimators, which may be used to determine the proper random dimensionality and the

584 order of approximation to balance the solution accuracy and computational efficiency.

585

Notation

C Covariance function

CV Coefficient of variation

c_j Coefficient of the jth PCE term

D Physical domain

d Order of polynomial chaos

f $f = \ln(K_s)$, log saturated hydraulic conductivity

g Sink/source term

H_B Prescribed pressure head on Dirichlet boundary segments

K Unsaturated hydraulic conductivity

K_s Saturated hydraulic conductivity

m Fitting parameter in the van Genuchten-Mualem model

N Random dimensionality (Number of KL terms retained to represent the mean-removed fields)

n Fitting parameter in the van Genuchten-Mualem model

P Number of PCE terms

Q Prescribed specific discharge on Neumann boundary segments

$\mathbf{q}$ Specific discharge

R_P Residual between the true solution and PCE approximation with P terms

$\mathbf{r}$ Separation vector

S_e Effective saturation

s $s = \ln(n-1)$

U Normal random input field ($U = f, \beta, s$)

609 V Lognormal random input field ($V = K_s, \alpha, (n-1)$)

610 $\mathbf{x}$ Cartesian coordinates in the physical domain

611 z Elevation

612 α Fitting parameter in the van Genuchten-Mualem model

613 β $\beta = \ln(\alpha)$

614 Φ_j Hermit polynomials

615 Γ_D Dirichlet boundary

616 Γ_N Neumann boundary

617 η_i Eigenvalues

618 λ Correlation length

619 θ_e Effective water content

620 θ_r Residual water content

621 θ_s Saturated water content

622 ρ Probability density

623 σ^2 Variance

624 Ω Probability space

625 ω Point in the probability space

626 $\boldsymbol{\xi}$ Collocation point

627 ξ_i Orthogonal standard Gaussian random variables

628 ψ Pressure head

629 $\hat{\psi}$ PCE approximation of the pressure head

630

631 ## Reference

632 Brooks, R.H., and A.T. Corey (1964), Hydraulic properties of porous media, *Hydrology*
633 *Paper 3*. Colorado State Univ., Fort Collins.

634 Courant, R., and D. Hilbert (1953), *Methods of Mathematical Physics*, Interscience, New
635 York.

636 Gardner, W.R. (1958), Some steady state solutions of unsaturated moisture flow equations
637 with application to evaporation from a water table, *Soil Sci. 85*, 228-232.

638 Ghanem, R., and P. Spanos (1991), *Stochastic Finite Element: A Spectral Approach*, Springer,
639 New York.

640 Jury, W.A. (1982), Simulation of solute transport using a transfer function model, *Water*
641 *Resour. Res. 18*, 363-368.

642 Li, H., and D. Zhang (2007), Probabilistic collocation method for flow in porous media:
643 Comparisons with other stochastic methods, *Water Resour. Res. 43*, W09409,
644 doi:10.1029/2006WR005673.

645 Lu, Z., and D. Zhang (2002), Stochastic analysis of transient flow in heterogeneous, variably
646 saturated porous media: the van Genuchten–Mualem constitutive model, *Vadose Zone J. 1*,
647 137-149.

648 Mantoglou, A. (1992), A theoretical approach for modeling unsaturated flow in spatially
649 variable soils: effective flow models in finite domains and nonstationarity, *Water Resour.*
650 *Res. 28(1)*, 251-267.

651 Mantoglou, A., L. Gelhar (1987), Stochastic modeling of large scale transient unsaturated
652 flow systems. *Water Resour. Res. 23(1)*, 37-46.

653 Russo, D. (1988), Determining soil hydraulic properties by parameter estimation: on the
654 selection of a model for the hydraulic properties, *Water Resour. Res. 24*(3), 453–459.
655 Russo, D. (1993), Stochastic modeling of macrodispersion for solute transport in a
656 heterogeneous unsaturated porous formation, *Water Resour. Res. 29*(2), 383–397.
657 Russo, D. (1995), On the velocity covariance and transport modeling in heterogeneous
658 anisotropic porous formations, 2, unsaturated flow, *Water Resour. Res. 31(1)*, 139-145.
659 Simunek, J., O. Wendroth, and M.T. van Genuchten (1998), Parameter estimation analysis of
660 the evaporation method for determining soil hydraulic properties, *Soil Sci. Soc. Am. J. 62*,
661 894-905.
662 Tatang, M. A., W. Pan, R. G. Prinn, and G. J. McRae (1997), An efficient method for
663 parametric uncertainty analysis of numerical geophysical models, *J. Geophy. Res. 102*
664 *(D18)*, 21925-21931.
665 van Genuchten, M.Th. (1980), A closed-form equation for predicting the hydraulic
666 conductivity of unsaturated soils, *Soil Sci. Soc. Am. J. 44*, 892–898.
667 Webster, M., M. A. Tatang, and G. J. Mcrae (1996), Application of the probabilistic
668 collocation method for an uncertainty analysis of a simple ocean model, MIT Joint Program
669 on the Science and Policy of Global Change Report Series No. 4, MIT.
670 Wiener, N. (1938), The homogeneous chaos, *Amer. J. Math., 60*, 897-936.
671 Yang, J., D. Zhang, and Z. Lu (2004), Stochastic analysis of saturated-unsaturated flow in
672 heterogeneous media by combining Karhunen-Loeve expansion and perturbation method, *J.*
673 *Hydrology, 294*, 18-38.
674 Yeh, T.-C., Gelhar, L.W., A.L. Gutjahr (1985a), Stochastic analysis of unsaturated flow in
675 heterogeneous soils, 1, statistically isotropic media, *Water Resour. Res. 21(4)*, 447-456.
676 Yeh, T.-C., Gelhar, L.W., A.L. Gutjahr (1985b), Stochastic analysis of unsaturated flow in

677 heterogeneous soils, 2, statistically anisotropic media with variable α, *Water Resour. Res.*
678 *21(4)*, 457-464.
679 Zhang, D., and C. L. Winter (1998), Nonstationary stochastic analysis of steady state flow
680 through variably saturated, heterogeneous media, *Water Resour. Res. 34*(5), 1091–1100.
681 Zhang, D. (1998), Numerical solutions to statistical moment equations of groundwater flow
682 in nonstationary, bounded heterogeneous media, *Water Resour. Res. 34(3)*, 529–538.
683 Zhang, D. (1999), Nonstationary stochastic analysis of transient unsaturated flow in randomly
684 heterogeneous media, *Water Resour. Res. 35(4)*, 1127–1141.
685 Zhang, D. (2002), *Stochastic Methods for Flow in Porous Media:Coping With Uncertainties*,
686 Academic, San Diego, CA.
687 Zyvoloski, G.A., B.A. Robinson, Z.V. Dash, and L.L. Trease (1997), Summary of the models
688 and methods for the FEHM application--A finite-element heat- and mass-transfer code,
689 LA-13307-MS, Los Alamos National Laboratory, Los Alamos, NM.

690

691 ## Table captions

692 Table 1. Summary of the parameters in all illustrative cases

693

694

695 Table 1. Summary of the parameters in all illustrative cases

Cases	$< K_s >$ (m/d)	CV_{Ks} (%)	λ_{fx} (m)	λ_{fy} (m)	$< \alpha >$ $(1/m)$	CV_{α} (%)	$\lambda_{\beta x}$ (m)	$\lambda_{\beta y}$ (m)	$< n >$	$CV_{(n-1)}$ (%)	λ_{sx} (m)	λ_{sy} (m)	Random dimensionality
1	1	100	1	1	2	0	NA	NA	1.4	0	NA	NA	20
2	1	50	1	1	2	0	NA	NA	1.4	0	NA	NA	20
3	1	100	0.5	0.5	2	0	NA	NA	1.4	0	NA	NA	20
4	1	100	0.5	0.5	2	0	NA	NA	1.4	0	NA	NA	30
5[a]	1	10	1	1	2	10	1	1	1.4	5	1	1	10+15+15=40
6[b]	1	10	1	1	2	10	1	1	1.4	5	1	1	15+15=30

696 [a] Case 5: f, β and s are uncorrelated;

697 [b] Case 6: f and β are perfectly correlated but are uncorrelated with s

698

699 ## Figure captions

700 Fig.1 Selected eigenfunctions for Case 1 ($\lambda_x = \lambda_y = 1.0m$)

701 Fig.2 Eigenvalues of the separate exponential covariance with different correlation scales

702 Fig.3 Comparison of results from MC and PCM methods for Case 1

703 ($CV_{Ks} = 100\%$, $\lambda_x = \lambda_y = 1.0m$) along the vertical central line

704 Fig.4 The mean and variance given by MC with different number of simulations, case 1

705 Fig.5 Contour maps of the pressure variance computed from MC simulations with a different

706 number of realizations (case1)

707 Fig.6 Comparison of pressure variance derived from PCM and MC, with different input

708 variance and different correlation length, cases 1, 2, 3 and 4

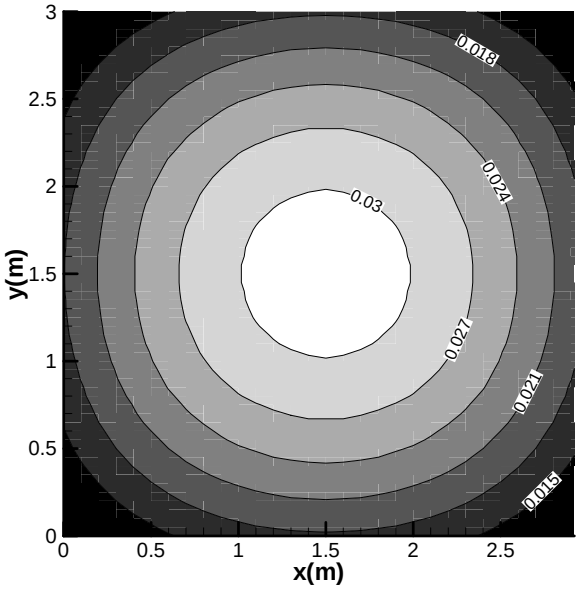
709 Fig.7 Statistics of the pressure head, with three uncorrelated input fields, case 5

710 Fig.8 Statistics of the pressure head, with perfectly correlated $f = \ln(Ks)$ and $\beta = \ln(\alpha)$, case 6

711

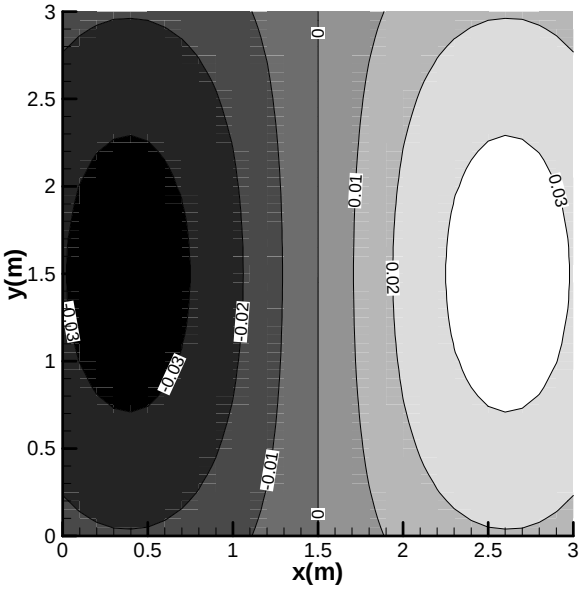
712 **Figures**

713

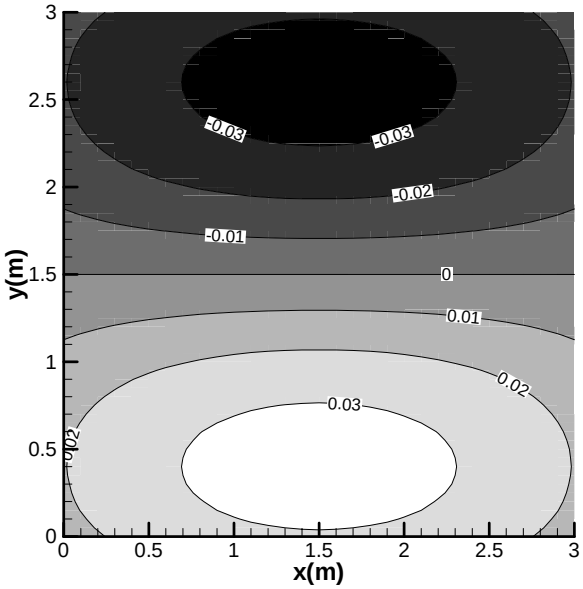


714

715 (a) The 1st eigenfunction

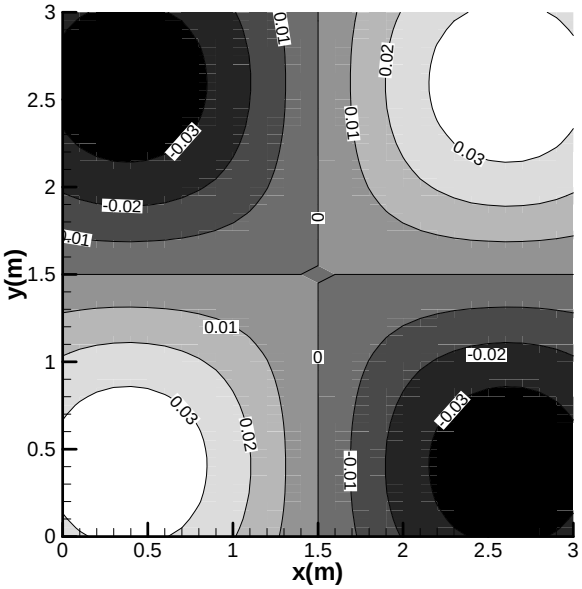


716 (b) The 2nd eigenfunction



717

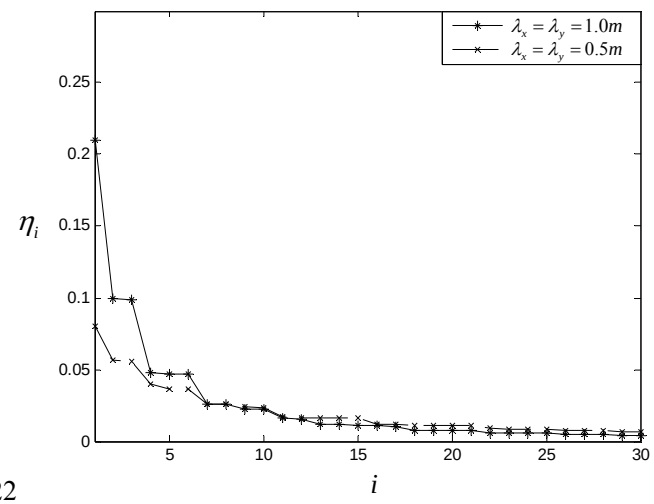
718 (c) The 3rd eigenfunction



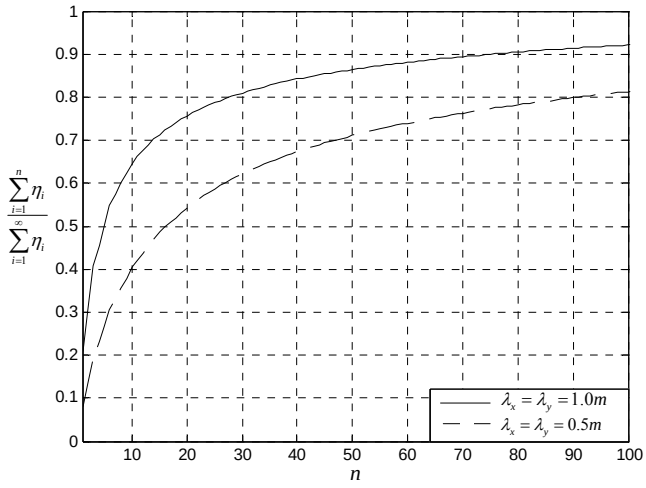
719 (d) The 6th eigenfunction

720 Fig. 1 Selected eigenfunctions for Case 1 ($\lambda_x = \lambda_y = 1.0m$)

721

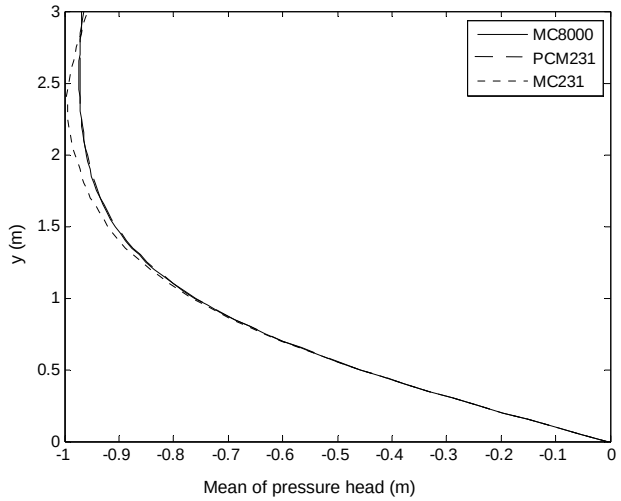


(a) Eigenvalues

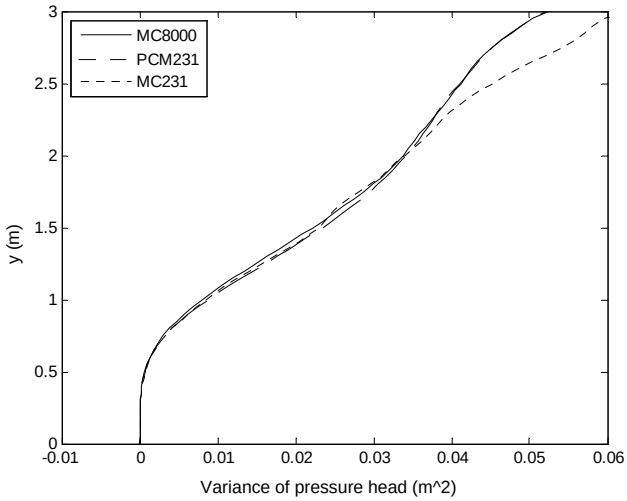


(b) Summation of eigenvalues

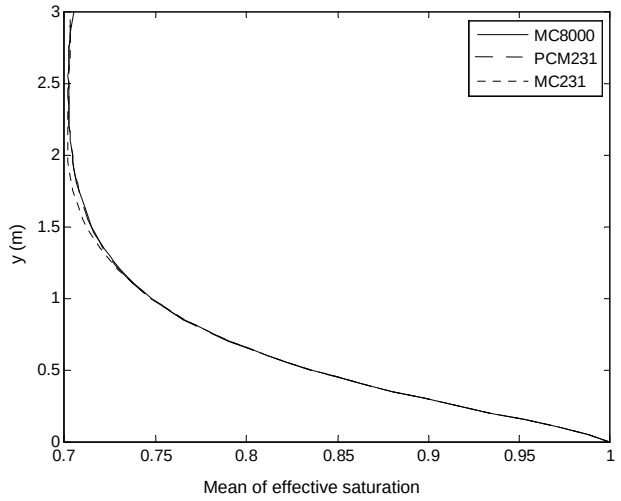
Fig. 2 Eigenvalues of separate exponential covariance with different correlation scales



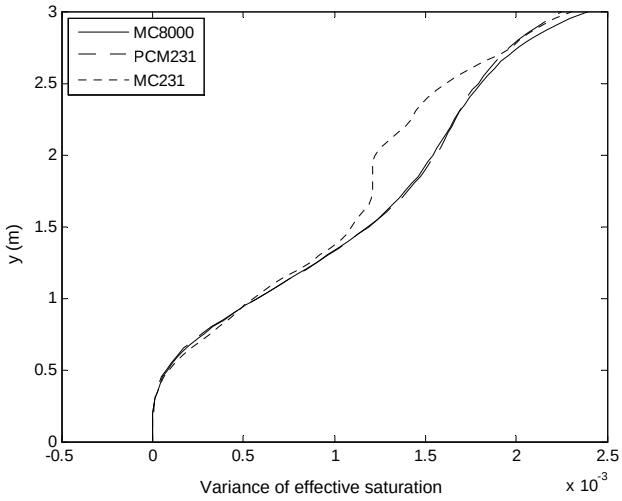
(a)



(b)

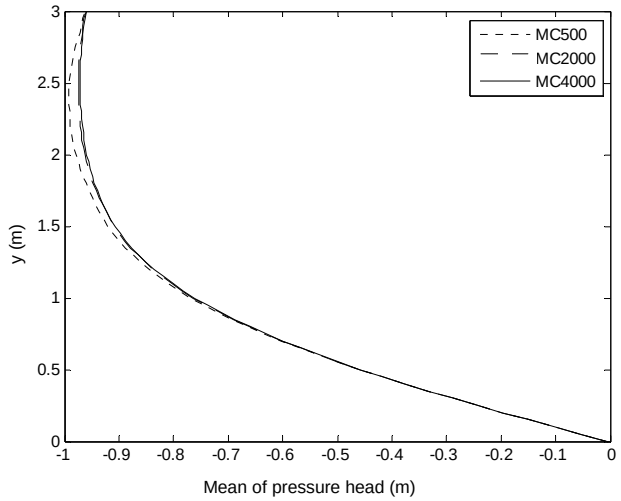


(c)

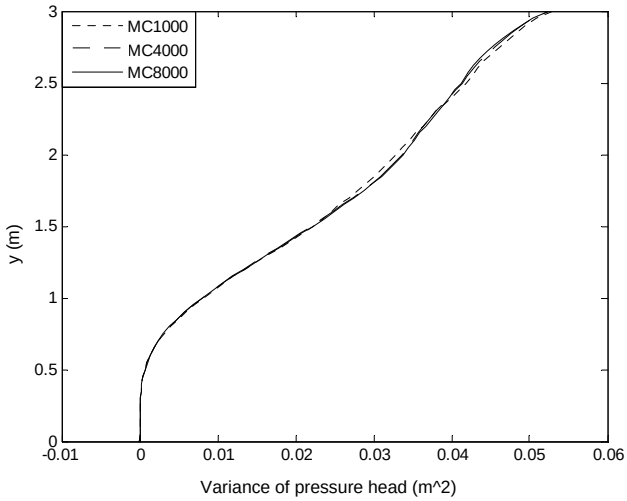


(d)

Fig.3 Comparison of results from MC and PCM methods for Case 1 ($CV_{Ks} = 100\%$, $\lambda_x = \lambda_y = 1.0m$) along the vertical central line.

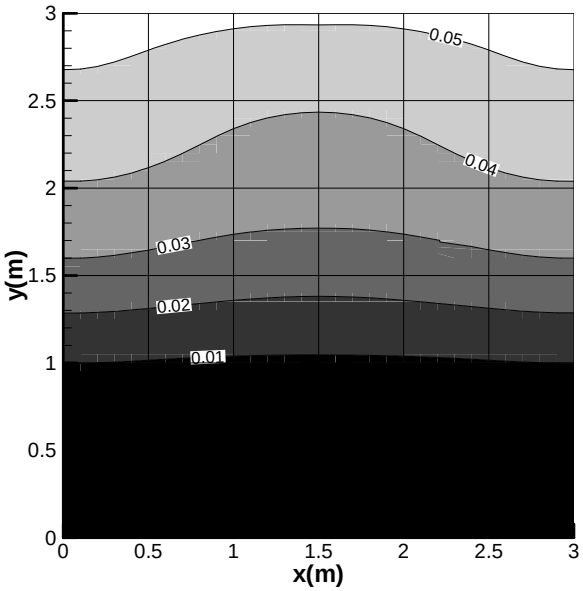


(a)



(b)

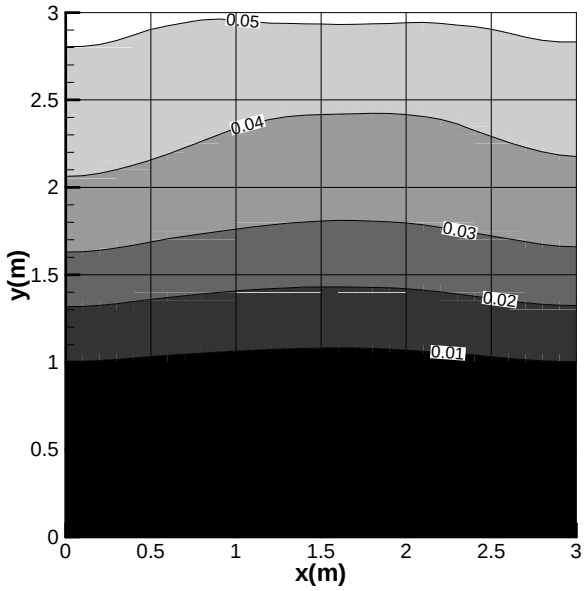
Fig. 4 The mean and variance given by MC with different number of simulations, case 1

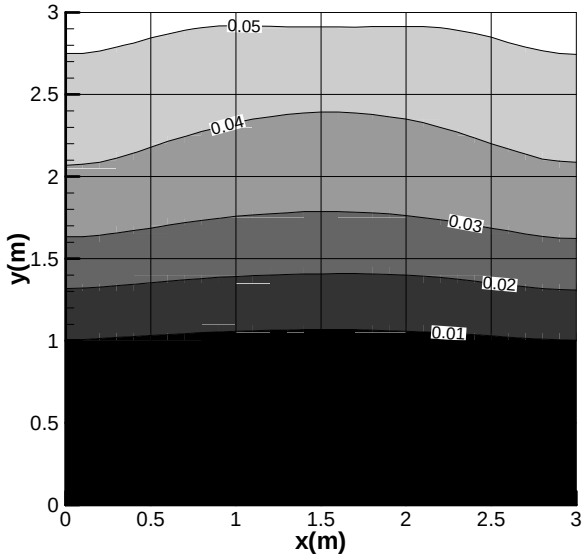


739

740 (a) σ_{ψ}^{2} (m^2), given by PCM, 231 simulations

741





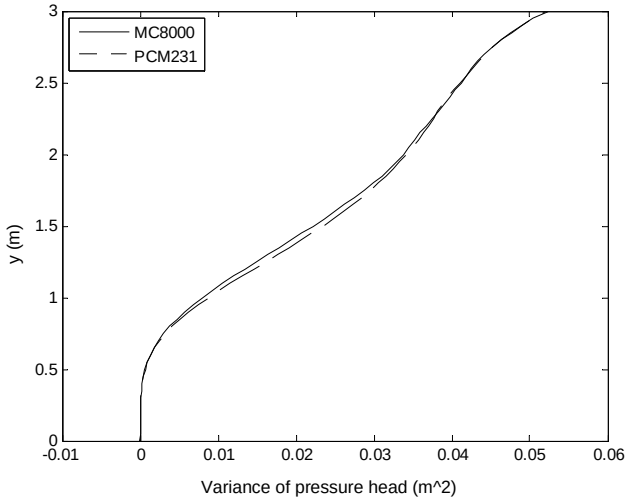
742

743 (b) σ_{ψ}^{2} (m^2), given by MC, 4000 simulations (c) σ_{ψ}^{2} (m^2), given by MC, 16000 simulations

744

745 Fig.5 Contour of the pressure variance in case 1

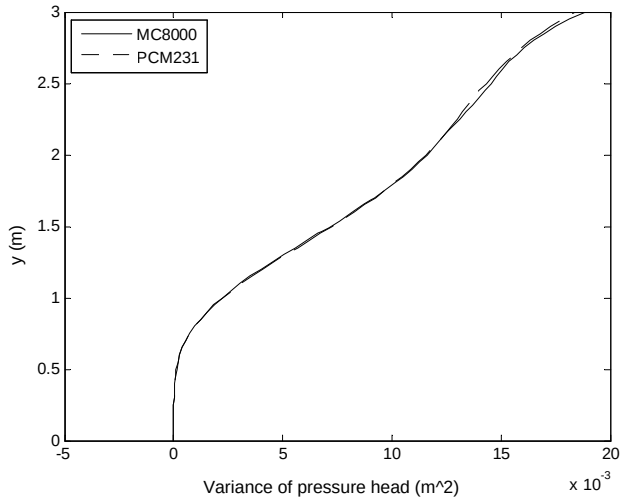
746



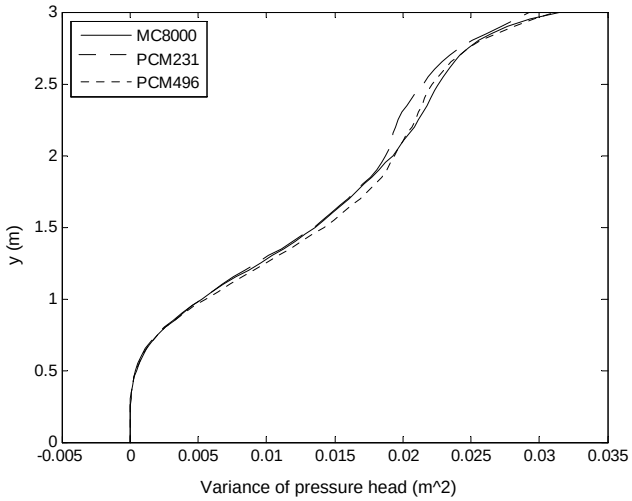
747

748 (a) case 1. $CV_{Ks} = 100\%$, $\lambda_x = \lambda_y = 1.0m$

749



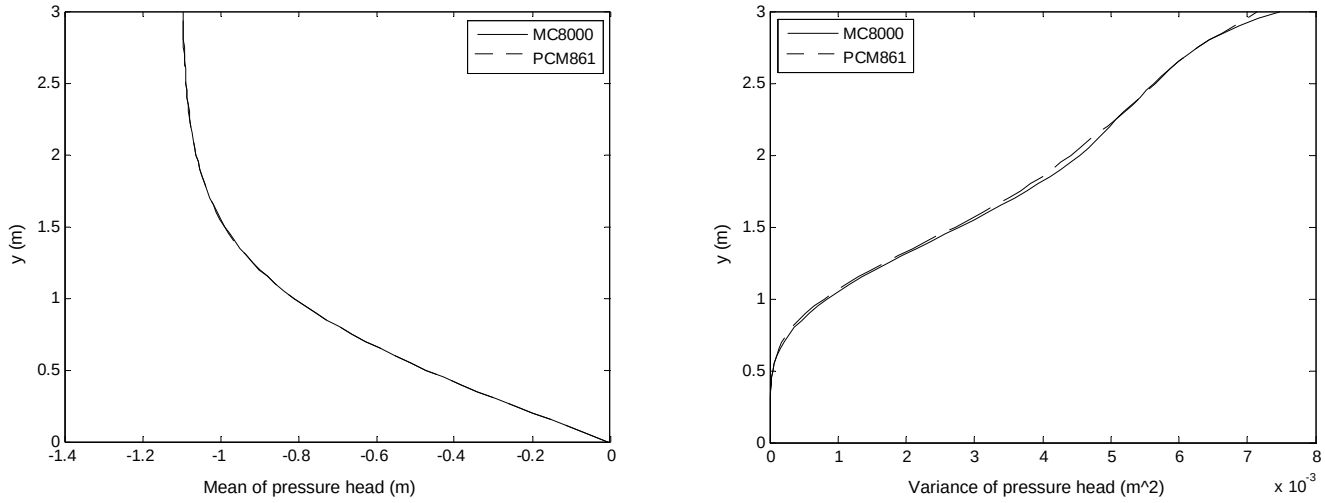
750



751 (b) case 2. $CV_{Ks} = 50\%$, $\lambda_x = \lambda_y = 1.0m$ (c) case 3 and 4 $CV_{Ks} = 100\%$, $\lambda_x = \lambda_y = 0.5m$

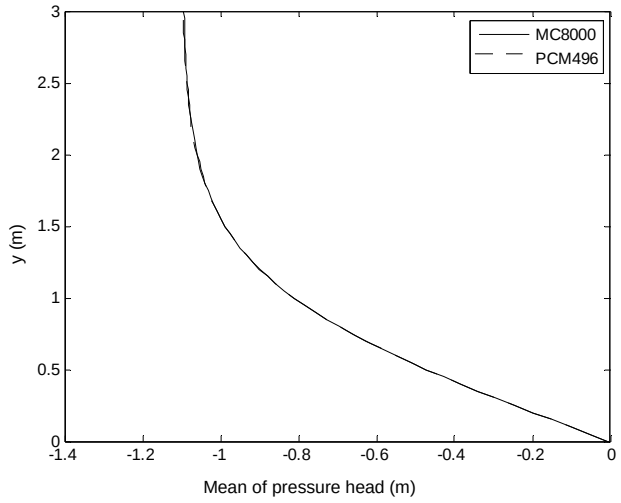
752 Fig.6 Comparison of pressure variance derived from PCM and MC, with different input variance and different
753 correlation length, cases 1, 2, 3 and 4

754

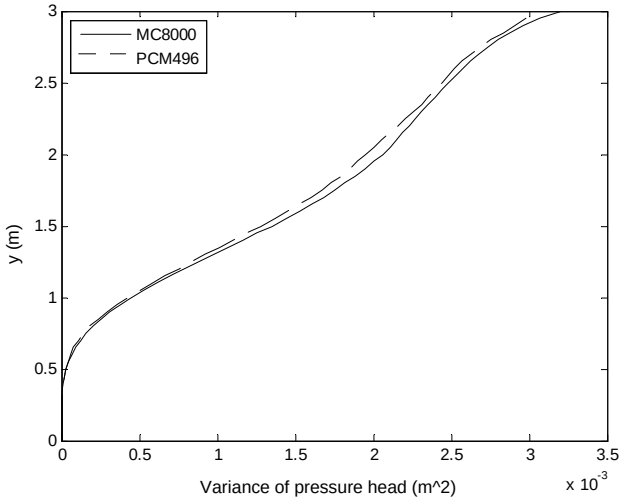


(a) (b)

Fig.7 Statistics of the pressure head, with three uncorrelated input fields (case 5).



(a)



(b)

Fig.8 Statistics of the pressure head, with perfectly correlated $f = \ln(Ks)$ and $\beta = \ln(\alpha)$, case 6.